

In [2]:

```
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

In [3]:

```
from transformers import AutoModelForCausalLM, AutoTokenizer
from transformers.generation import GenerationConfig
import warnings
warnings.filterwarnings("ignore")
import textwrap
import torch
import time
```

```
!pip install tiktoken
!pip install tiktoken transformers_stream_generator einops optimum auto-gptq
```

```
from transformers import pipeline
```

```
tokenizer = AutoTokenizer.from_pretrained("Qwen/Qwen-1.8B-Chat", trust_remote_code=True)
pipe = pipeline("text-generation", model="Qwen/Qwen-1.8B-Chat", device_map="auto", trust_remote_code=True)
```

Requirement already satisfied: tiktoken in /usr/local/lib/python3.10/dist-packages (0.6.0)

Requirement already satisfied: regex>=2022.1.18 in /usr/local/lib/python3.10/dist-packages (from tiktoken) (2023.12.25)

Requirement already satisfied: requests>=2.26.0 in /usr/local/lib/python3.10/dist-packages (from tiktoken) (2.31.0)

Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken) (2.0.7)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken) (2024.2.2)

Requirement already satisfied: tiktoken in /usr/local/lib/python3.10/dist-packages (0.6.0)

Requirement already satisfied: transformers\_stream\_generator in /usr/local/lib/python3.10/dist-packages (0.0.5)

Requirement already satisfied: einops in /usr/local/lib/python3.10/dist-packages (0.8.0)

Requirement already satisfied: optimum in /usr/local/lib/python3.10/dist-packages (1.19.1)

Requirement already satisfied: auto-gptq in /usr/local/lib/python3.10/dist-packages (0.7.1)

Requirement already satisfied: regex>=2022.1.18 in /usr/local/lib/python3.10/dist-packages (from tiktoken) (2023.12.25)

Requirement already satisfied: requests>=2.26.0 in /usr/local/lib/python3.10/dist-packages (from tiktoken) (2.31.0)

Requirement already satisfied: transformers>=4.26.1 in /usr/local/lib/python3.10/dist-packages (from transformers\_stream\_generator) (4.40.1)

Requirement already satisfied: coloredlogs in /usr/local/lib/python3.10/dist-packages (from optimum) (15.0.1)

Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from optimum) (1.12)

Requirement already satisfied: torch>=1.11 in /usr/local/lib/python3.10/dist-packages (from optimum) (2.2.1+cu121)

Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from optimum) (24.0)

Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from optimum) (1.25.2)

Requirement already satisfied: huggingface-hub>=0.8.0 in /usr/local/lib/python3.10/dist-packages (from optimum) (0.22.2)

Requirement already satisfied: datasets in /usr/local/lib/python3.10/dist-packages (from

optimum) (2.19.0)  
Requirement already satisfied: accelerate>=0.26.0 in /usr/local/lib/python3.10/dist-packages (from auto-gptq) (0.29.3)  
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.10/dist-packages (from auto-gptq) (0.1.99)  
Requirement already satisfied: rouge in /usr/local/lib/python3.10/dist-packages (from auto-gptq) (1.0.1)  
Requirement already satisfied: gekko in /usr/local/lib/python3.10/dist-packages (from auto-gptq) (1.1.1)  
Requirement already satisfied: safetensors in /usr/local/lib/python3.10/dist-packages (from auto-gptq) (0.4.3)  
Requirement already satisfied: peft>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from auto-gptq) (0.10.0)  
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from auto-gptq) (4.66.2)  
Requirement already satisfied: psutil in /usr/local/lib/python3.10/dist-packages (from accelerate>=0.26.0->auto-gptq) (5.9.5)  
Requirement already satisfied: pyyaml in /usr/local/lib/python3.10/dist-packages (from accelerate>=0.26.0->auto-gptq) (6.0.1)  
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.8.0->optimum) (3.13.4)  
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.8.0->optimum) (2023.6.0)  
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.8.0->optimum) (4.11.0)  
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken) (3.3.2)  
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken) (3.7)  
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken) (2.0.7)  
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.26.0->tiktoken) (2024.2.2)  
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (3.3)  
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (3.1.3)  
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (12.1.105)  
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (12.1.105)  
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (12.1.105)  
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (8.9.2.26)  
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (12.1.3.1)  
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (11.0.2.54)  
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (10.3.2.106)  
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (11.4.5.107)  
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (12.1.0.106)  
Requirement already satisfied: nvidia-nccl-cu12==2.19.3 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (2.19.3)  
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (12.1.105)  
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.11->optimum) (2.2.0)  
Requirement already satisfied: nvidia-nvjitlink-cu12 in /usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-cu12==11.4.5.107->torch>=1.11->optimum) (12.4.127)  
Requirement already satisfied: tokenizers<0.20,>=0.19 in /usr/local/lib/python3.10/dist-packages (from transformers>=4.26.1->transformers\_stream\_generator) (0.19.1)  
Requirement already satisfied: protobuf in /usr/local/lib/python3.10/dist-packages (from transformers>=4.26.1->transformers\_stream\_generator) (3.20.3)  
Requirement already satisfied: humanfriendly>=9.1 in /usr/local/lib/python3.10/dist-packages (from coloredlogs->optimum) (10.0)  
Requirement already satisfied: pyarrow>=12.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets->optimum) (14.0.2)  
Requirement already satisfied: pyarrow-hotfix in /usr/local/lib/python3.10/dist-packages

(from datasets->optimum) (0.6)  
Requirement already satisfied: dill<0.3.9,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from datasets->optimum) (0.3.8)  
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets->optimum) (2.0.3)  
Requirement already satisfied: xxhash in /usr/local/lib/python3.10/dist-packages (from datasets->optimum) (3.4.1)  
Requirement already satisfied: multiprocessing in /usr/local/lib/python3.10/dist-packages (from datasets->optimum) (0.70.16)  
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets->optimum) (3.9.5)  
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from rouge->auto-gptq) (1.16.0)  
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->optimum) (1.3.0)  
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets->optimum) (1.3.1)  
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets->optimum) (23.2.0)  
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets->optimum) (1.4.1)  
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets->optimum) (6.0.5)  
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets->optimum) (1.9.4)  
Requirement already satisfied: async-timeout<5.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets->optimum) (4.0.3)  
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch>=1.11->optimum) (2.1.5)  
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets->optimum) (2.8.2)  
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets->optimum) (2023.4)  
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets->optimum) (2024.1)

In [4]:

```
def get_completion(prompt):  
    messages = [{  
        "role": "user",  
        "content": prompt }]  
    prompt2 = pipe.tokenizer.apply_chat_template(messages, tokenize=False, add_generation_prompt=True)  
    outputs = pipe(prompt2, max_new_tokens=150, do_sample=True, temperature=0.7, top_k=50, top_p=0.95)  
    return outputs[0]["generated_text"]
```

## Main Data

In [5]:

```
# Define the question  
question = "What are your refund policies?"  
  
# Read the content from the output file  
with open('/content/drive/MyDrive/Colab Notebooks/outputMain.txt', 'r') as file:  
    content = file.read().strip()  
  
# Construct the prompt to pass to the LLM  
prompt = f"""  
Based on the detailed information below, please provide a comprehensive answer to the customer's query:  
  
{content}  
  
Question: {question}  
  
Your response should comprehensively integrate the information provided, accurately refle
```

```
cting the key points and offering a clear, well-structured answer.
"""
```

```
# Generate the response using the LLM
start_time = time.time()
response = get_completion(prompt)
end_time = time.time()
print(response)
print("Time taken: ", end_time - start_time)
```

No chat template is defined for this tokenizer - using a default chat template that implements the ChatML format (without BOS/EOS tokens!). If the default is not appropriate for your model, please set `tokenizer.chat\_template` to an appropriate template. See [https://huggingface.co/docs/transformers/main/chat\\_templating](https://huggingface.co/docs/transformers/main/chat_templating) for more information.

```
<|im_start|>user
```

Based on the detailed information below, please provide a comprehensive answer to the customer's query:

If the item received is damaged, defective, incorrect, or incomplete, a refund will be issued based on Daraz's assessment. Note: For device-related issues after usage or expiration of return policy period, please check if the item is covered under Seller or Brand Warranty. Items that are non-returnable: Any custom-made items" What is the cancellation and return policy for Buy More, Save More orders? Partial cancellations are not allowed but partial returns are allowed on Buy More, Save More orders i. e. you will not be able to cancel 1 or 2 products in your order and would instead have to cancel the entire order. But if you decide to apply for a return, you can return certain items out of your total Buy More, Save More order. If a certain number of products from a Buy More, Save More order are eligible for a return, partial refund will be issued to the customer after these items have been returned by the customer & quality checked by Daraz. Partial refunds are calculated on item level of each bundle. In case you have received your order but are not satisfied with your purchase, you can submit a return request within 14 days from the day it is delivered to you. Decision on returns and refunds taken by Daraz will be final in its sole discretion. In case a customer tries to circumvent the terms and conditions and abuse subsidies by creating new accounts, they will be unable to place their order. "

Question: What are your refund policies?

Your response should comprehensively integrate the information provided, accurately reflecting the key points and offering a clear, well-structured answer.

```
<|im_end|>
```

```
<|im_start|>assistant
```

Thank you for your question. Daraz offers a flexible refund policy that allows customers to request a full or partial refund based on their assessment of the item received. The policy also includes provisions for partial cancellations and returns, as well as guidelines on what items may be eligible for a return.

Regarding the refund policy for Buy More, Save More orders, partial cancellations are not allowed. However, partial refunds are available on certain items where a certain number of products from the order meet the criteria for a return. Partial refunds are calculated on an item-level basis, meaning that only a certain percentage of the total cost of the order will be refunded. In most cases, partial refunds will be issued to the customer after the items have been returned by the customer

Time taken: 161.6787304878235

## Seller Data

In [6]:

```
# Define the question
question = "Can I make product bundles on Daraz?"

# Read the content from the output file
with open('/content/drive/MyDrive/Colab Notebooks/outputSeller.txt', 'r') as file:
    content = file.read().strip()

# Construct the prompt to pass to the LLM
prompt = f"""
```

Based on the detailed information below, please provide a comprehensive answer to the customer's query:

```
{content}
```

```
Question: {question}
```

Your response should comprehensively integrate the information provided, accurately reflecting the key points and offering a clear, well-structured answer.

```
# Generate the response using the LLM
start_time = time.time()
response = get_completion(prompt)
end_time = time.time()
print(response)
print("Time taken: ", end_time - start_time)
```

```
<|im_start|>user
```

Based on the detailed information below, please provide a comprehensive answer to the customer's query:

You must deactivate one bundle feature for that SKU in order to reactivate another. Bundles: Buy 1 Get 1 Free A common form of sales promotion also known as two for one whilst the cost per item is proportionately cheaper than if bought on its own, it is not actually half price. Bundles: Combo A promotion where different products are sold together as a set with one or more items at discounted prices Products with high views and sales mix with less popular products. Minimum 5 custom products should be there if a seller is selling only custom made products What are the Benefits of Product Customization? Match Customer Demand Increase sales and competitive edge Customer engagement Build brand loyalty Gain market research" Why should I create bundles? Sellers can create bundles to differentiate their deals from competitors, especially in high-priced competitive categories It's a great opportunity to sell low-cost items as part of a bundle Bundles allow sellers to clear their inventory of whatever products they choose Additionally, bundles are a foolproof way to attract customers and increase conversion rate, sales and enhance overall performance. Sellers sell a package or set of goods for a lower price than they would charge if the customer bought all of them separately. Minimum 5 custom products should be there if a seller is selling only custom made products What are the Benefits of Product Customization? Match Customer Demand Increase sales and competitive edge Customer engagement Build brand loyalty Gain market research" What are the types of Product Bundling? Minimum 5 custom products should be there if a seller is selling only custom made products What are the Benefits of Product Customization?

```
Question: Can I make product bundles on Daraz?
```

Your response should comprehensively integrate the information provided, accurately reflecting the key points and offering a clear, well-structured answer.

```
<|im_end|>
```

```
<|im_start|>assistant
```

Yes, you can create product bundles on Daraz. In fact, Daraz is an e-commerce platform that offers various product bundling options to help sellers increase sales and compete in the high-priced competitive category. To create a bundle on Daraz, you will need to meet certain minimum requirements such as having at least 5 custom-made products and being selling only custom-made products.

Product bundling offers several benefits to sellers, including:

1. Differentiating your deals from competitors: By creating bundles that offer a discount compared to buying each item separately, you can differentiate your deals from competitors and attract customers who may be looking for a better deal.

2. Increasing sales: Bundle deals can help you increase sales by encouraging customers to purchase multiple items instead of

Time taken: 153.29610395431519

## Product Data

In [8]:

```

# Define the question
question = "Can you list watches between Rs. 1000 and Rs. 2000?"

# Read the content from the output file
with open('/content/drive/MyDrive/Colab Notebooks/outputProducts.txt', 'r') as file:
    content = file.read().strip()

# Construct the prompt to pass to the LLM
prompt = f"""
Customer Question: {question}
Product List:

{content}

Based on the above product list, please provide a helpful and engaging answer that encourages the customer to consider the options and possibly make a purchase. Highlight the key features and benefits of the products listed.
"""

# Generate the response using the LLM
start_time = time.time()
response = get_completion(prompt)
end_time = time.time()
print(response)
print("Time taken: ", end_time - start_time)

```

<|im\_start|>user

Customer Question: Can you list watches between Rs. 1000 and Rs. 2000?  
Product List:

Product 04: Product Name = Men Quartz Watch for Waterproof Vacuum Plating Strap with Car Wheel Rim Hub Design Sport Automatic Wristwatch (Does not rotate), Product Category = Watches Sunglasses Jewellery/Watches/Men/Fashion, Brand Name = No Brand, Seller Name = D&K Online, URL = <https://www.daraz.pk/products/-i411043073-s1964553711.html?search=1>, Price Details = Original: Rs. 3999, Discounted: Rs. 1768 | Original: Rs. 3999, Discounted: Rs. 1869 | Original: Rs. 3999, Discounted: Rs. 1769 | Original: Rs. 3999, Discounted: Rs. 1768 | Original: Rs. 3999, Discounted: Rs. 1868, Positive Seller Ratings = 82%, Ship on Time = 100%, Return Policy = 14 days free & easy return (Change of mind is not applicable)  
Product 10: Product Name = 2022 trend full-automatic quartz movement men's watch wheel no n mechanical watch wrist watch fashion men's watch, Product Category = Watches Sunglasses Jewellery/Watches/Men/Fashion, Brand Name = No Brand, Seller Name = lamgool Electronic, URL = <https://www.daraz.pk/products/2022-i366103747-s1819200832.html?search=1>, Price Details = Original: Rs. 1836, Discounted: Rs. 1410, Positive Seller Ratings = 81%, Ship on Time = 100%, Return Policy = 14 days free & easy return (Change of mind is not applicable)  
Product 17: Product Name = Qiunery YESURPRISE Unisex Antique Case Vintage Brass Rib Chain Quartz Pocket Watch Train, Product Category = Watches Sunglasses Jewellery/Watches/Men/Fashion, Brand Name = No Brand, Seller Name = Qiunery, URL = <https://www.daraz.pk/products/qiunery-yesur-i202992571-s1402460810.html?search=1>, Price Details = Original: Rs. 1790, Discounted: Rs. 1250, Positive Seller Ratings = 82%, Ship on Time = 100%, Return Policy = 14 days free & easy return (Change of mind is not applicable)

Based on the above product list, please provide a helpful and engaging answer that encourages the customer to consider the options and possibly make a purchase. Highlight the key features and benefits of the products listed.

<|im\_end|>

<|im\_start|>assistant

Hello! Thank you for your interest in our watches. We have some great options available that should suit your needs and budget.

For those looking for a more affordable option, we recommend the Men Quartz Watch for Waterproof Vacuum Plating strap with Car Wheel Rim Hub Design Sport Automatic Wristwatch. It has a sporty design and automatic movement, making it perfect for everyday wear or as a gift for a friend. The price is \$1768, but there are also discounts available for customers who purchase before a certain date.

If you're interested in a luxury experience, we suggest the 2022 trend full-automatic quartz movement men's watch. This watch comes with an antique case and vintage brass rib chain, providing a unique and stylish

Time taken: 182.77021598815918